Model Fusion:
A New Approach To Processing Heterogenous Data

Omar Ochoa
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
1. Need to Combine Data from Different Sources

- In many areas of science and engineering, we have different sources of data.
- For example, in geophysics, there are many sources of data for Earth models:
  - first-arrival passive seismic data (from the actual earthquakes);
  - first-arrival active seismic data (from the seismic experiments);
  - gravity data; and
  - surface waves.
2. Need to Combine Data (cont-d)

- Datasets coming from different sources provide complimentary information.

- *Example:* different geophysical datasets contain different information on earth structure.

- In general:
  - some of the datasets provide better accuracy and/or spatial resolution in some spatial areas;
  - other datasets provide a better accuracy and/or spatial resolution in other areas or depths.

- *Example:*
  - gravity measurements have (relatively) low spatial resolution;
  - a seismic data point comes from a narrow trajectory of a seismic signal – so spatial resolution is higher.
3. Joint Inversion: An Ideal Future Approach

• *At present:* each of the datasets is often processed separately.

• *It is desirable:* to data from different datasets.

• *Ideal approach:* use all the datasets to produce a single model.

• *Problem:* in many areas, there are no efficient algorithms for simultaneously processing all the datasets.

• *Challenge:* designing joint inversion techniques is an important theoretical and practical challenge.
4. Data Fusion: Case of Interval Uncertainty

- In some practical situations, the value \( x \) is known with interval uncertainty.
- This happens, e.g., when we only know the upper bound \( \Delta^{(i)} \) on each estimation error \( \Delta x^{(i)} \): \( |\Delta x^{(i)}| \leq \Delta^{(i)} \).
- In this case, we can conclude that \( |x - \tilde{x}^{(i)}| \leq \Delta^{(i)} \), i.e., that \( x \in x^{(i)} \overset{\text{def}}{=} [\tilde{x}^{(i)} - \Delta^{(i)}, \tilde{x}^{(i)} + \Delta^{(i)}] \).
- Based on each estimate \( \tilde{x}^{(i)} \), we know that the actual value \( x \) belongs to the interval \( x^{(i)} \).
- Thus, we know that the (unknown) actual value \( x \) belongs to the intersection of these intervals:

\[
x \overset{\text{def}}{=} \bigcap_{i=1}^{n} x^{(i)} = [\max(\tilde{x}^{(i)} - \Delta^{(i)}), \min(\tilde{x}^{(i)} + \Delta^{(i)})].
\]
5. Proposed Solution – Model Fusion: Main Idea

- **Reminder:** joint inversion methods are still being developed.

- **Practical solution:** to fuse the models coming from different datasets.

- **Simplest case – data fusion, probabilistic uncertainty:**
  - we have several estimates $\tilde{x}^{(1)}, \ldots, \tilde{x}^{(n)}$ of the same quantity $x$.
  - each estimation error $\Delta x^{(i)} \equiv \tilde{x}^{(i)} - x$ is normally distributed with 0 mean and known st. dev. $\sigma^{(i)}$;
  - Least Squares: find $x$ that minimizes \( \sum_{i=1}^{n} \frac{(\tilde{x}^{(i)} - x)^2}{2 \cdot (\sigma^{(i)})^2} \);

- solution: $x = \frac{\sum_{i=1}^{n} \tilde{x}^{(i)} \cdot (\sigma^{(i)})^{-2}}{\sum_{i=1}^{n} (\sigma^{(i)})^{-2}}$. 
6. Towards Formulation of a Problem

- *What is given:*
  - we have high spatial resolution estimates $\tilde{x}_1, \ldots, \tilde{x}_n$ of the values $x_1, \ldots, x_n$ in several small cells;
  - we also have low spatial resolution estimates $\tilde{X}_j$ for the weighted averages
    \[
    X_j = \sum_{i=1}^{n} w_{j,i} \cdot x_i.
    \]

- *Objective:* based on the estimates $\tilde{x}_i$ and $\tilde{X}_j$, we must provide more accurate estimates for $x_i$.

- *Geophysical example:* we are interested in the densities $x_i$. 
7. Model Fusion: Case of Probabilistic Uncertainty

We take into account several different types of approximate equalities:

- Each high spatial resolution value $\tilde{x}_i$ is approximately equal to the actual value $x_i$, w/known accuracy $\sigma_{h,i}$:
  \[
  \tilde{x}_i \approx x_i.
  \]

- Each lower spatial resolution value $\tilde{X}_j$ is approximately equal to the weighted average, w/known accuracy $\sigma_{l,j}$:
  \[
  \tilde{X}_j \approx \sum_i w_{j,i} \cdot x_i.
  \]

- We usually have a prior knowledge $x_{pr,i}$ of the values $x_i$, with accuracy $\sigma_{pr,i}$: $x_i \approx x_{pr,i}$.

- Also, each lower spatial resolution value $\tilde{X}_j$ is $\approx$ the value within each of the smaller cells:
  \[
  \tilde{X}_j \approx x_i(l,j).
  \]
8. Case of Probabilistic Uncertainty: Details

- Each lower spatial resolution value $\tilde{X}_j$ is approximately equal to the value within each of the smaller cells:
  $$\tilde{X}_j \approx x_i(l,j).$$

- The accuracy of $\tilde{X}_j \approx x_i(l,j)$ corresponds to the (empirical) standard deviation:
  $$\sigma^2_{e,j} \overset{\text{def}}{=} \frac{1}{k_j} \cdot \sum_{l=1}^{k_j} \left( \tilde{x}_i(l,j) - E_j \right)^2,$$

where
  $$E_j \overset{\text{def}}{=} \frac{1}{k_j} \cdot \sum_{l=1}^{k_j} \tilde{x}_i(l,j).$$
9. Model Fusion: Least Squares Approach

- **Main idea:** use the Least Squares technique to combine the approximate equalities.
- We find the desired combined values \( x_i \) by minimizing the corresponding sum of weighted squared differences:

\[
\sum_{i=1}^{n} \frac{(x_i - \tilde{x}_i)^2}{\sigma^2_{h,i}} + \sum_{j=1}^{m} \frac{1}{\sigma^2_{l,j}} \cdot \left( \tilde{X}_j - \sum_{i=1}^{n} w_{j,i} \cdot x_i \right)^2 + \\
\sum_{i=1}^{n} \frac{(x_i - x_{pr,i})^2}{\sigma^2_{pr,i}} + \sum_{j=1}^{m} \sum_{l=1}^{k_j} \frac{(\tilde{X}_j - x_{i(l,j)})^2}{\sigma^2_{e,j}}.
\]
10. Model Fusion: Solution

- To find a minimum of an expression, we:
  - differentiate it with respect to the unknowns, and
  - equate derivatives to 0.

- Differentiation with respect to $x_i$ leads to the following system of linear equations:

\[
\frac{1}{\sigma_{h,i}^2} \cdot (x_i - \tilde{x}_i) + \sum_{j:j\ni i} \frac{1}{\sigma_{l,j}^2} \cdot w_{j,i} \cdot \left(\sum_{i'=1}^{n} w_{j,i'} \cdot x_{i'} - \tilde{X}_j \right) + \\
\frac{1}{\sigma_{pr,i}^2} \cdot (x_i - x_{pr,i}) + \sum_{j:j\ni i} \frac{1}{\sigma_{e,j}^2} \cdot (x_i - \tilde{X}_j) = 0,
\]

where $j \ni i$ means that the $j$-th low spatial resolution estimate covers $i$-th cell.

- **Idea:** fuse each high spatial resolution estimate $\tilde{x}_i$ with a prior estimate $x_{pr,i}$.

- **Detail:** instead of
  \[
  \frac{1}{\sigma^2_{h,i}} \cdot (x_i - \tilde{x}_i) + \frac{1}{\sigma^2_{pr,i}} \cdot (x_i - x_{pr,i}),
  \]
  we have a single term
  \[
  \sigma^{-2}_{f,i} \cdot (x_i - x_{f,i}),
  \]
  where
  \[
  x_{f,i} \overset{\text{def}}{=} \frac{\tilde{x}_i \cdot \sigma^{-2}_{h,i} + x_{pr,i} \cdot \sigma^{-2}_{pr,i}}{\sigma^{-2}_{h,i} + \sigma^{-2}_{pr,i}}, \quad \sigma^{-2}_{f,i} \overset{\text{def}}{=} \sigma^{-2}_{h,i} + \sigma^{-2}_{pr,i}.
  \]

- **Resulting simplified equations:**
  \[
  \sigma^{-2}_{f,i} \cdot (x_i - x_{f,i}) + \sum_{j: j \ni i} \frac{1}{\sigma^2_{l,j}} \cdot w_{j,i} \cdot \left( \sum_{i'=1}^{n} w_{j,i'} \cdot x_{i'} - \tilde{X}_j \right) + \sum_{j: j \ni i} \frac{1}{\sigma^2_{e,j}} \cdot (x_i - \tilde{X}_j) = 0.
  \]
12. Case of a Single Low Spatial Resolution Estimate

- **Simplest case:** we have exactly one low spatial resolution estimate $\tilde{X}_1$.

- **In general:** we only have high spatial resolution estimates for *some* of the cells.

- **In geosciences:** such a situation is typical: e.g.,
  - we have a low spatial resolution gravity estimates which cover a huge area in depth, and
  - we have high spatial resolution seismic estimates which only cover depths above the Moho.

- **For convenience:** let us number the cells for which we have high spatial resolution estimates first.

- Let $h$ denote the total number of such cells.
13. Case of a Single Low Spatial Resolution Estimate: Simplified Algorithm

First, we compute the auxiliary value

\[
\mu \overset{\text{def}}{=} \frac{1}{\sigma^2_{l,1}} \cdot \left( \sum_{i'} w_{1,i'} \cdot x_{i'} - \tilde{X}_1 \right)
\]

as \(\mu = \frac{N}{D}\), where

\[
N = \sum_{i=1}^{h} \frac{w_{1,i} \cdot (x_{f,i} - \tilde{X}_1)}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}},
\]

and

\[
D = \sigma^2_{l,1} + \sum_{i=1}^{h} \frac{w_{1,i}^2 \cdot \sigma_{f,i}^2}{1 + \frac{\sigma_{f,i}^2}{\sigma_{e,1}^2}} + \left( \sum_{i=h+1}^{n} w_{1,i}^2 \right) \cdot \sigma_{e,1}^2.
\]
14. Case of a Single Low Spatial Resolution Estimate: Simplified Algorithm (cont-d)

- Once we know $\mu$, we compute the desired estimates for $x_i$, $i = 1, \ldots, h$, as

$$x_i = \frac{x_{f,i}}{\sigma^2_{e,1}} - \frac{w_{1,i} \cdot \sigma^2_{f,i}}{1 + \frac{\sigma^2_{f,i}}{\sigma^2_{e,1}}} \cdot \mu + \bar{X}_1 \cdot \frac{\sigma^2_{f,i}}{1 + \frac{\sigma^2_{f,i}}{\sigma^2_{e,1}}}.$$  

- We also compute estimates $x_i$ for $i = h + 1, \ldots, n$, as

$$x_i = \bar{X}_1 - w_{1,i} \cdot \sigma^2_{e,1} \cdot \mu.$$
15. Numerical Example: Description

- **Objective:** to illustrate the above formulas.
- **Idea:** consider the simplest possible case, when we have
  - exactly one low spatial resolution estimate \( \tilde{X}_1 \)
  - that covers all \( n \) cells,

and when:

- all the weights are all equal \( w_{1,i} = 1/n \);
- there is a high spatial resolution estimate corresponding to each cell \( (h = n) \);
- all high spatial resolution estimates have the same accuracy \( \sigma_{h,i} = \sigma_h \);
- \( \sigma_{l,1} \ll \sigma_h \), so \( \sigma_{l,1} \approx 0 \); and
- there is no prior information, so \( \sigma_{pr,i} = \infty \) and thus, \( x_{f,i} = \tilde{x}_i \) and \( \sigma_{f,i} = \sigma_h \).
16. Additional Simplification

- **In general:** there are cells for which there are no high spatial resolution estimates.

- **How to deal with these cells:** we added a heuristic rule that
  
  - each lower spatial resolution value is approximately equal to the value within each of the constituent cells,
  
  - with the accuracy corresponding to the (empirical) standard deviation $\sigma_{e,j}$.

- **In our simplified example:** we have high spatial resolution estimate in each cell.

- So, there is no need for this heuristic rule.

- The corresponding heuristic terms in the least squares approach are proportional to $\frac{1}{\sigma_{e,1}^2}$, so we take $\sigma_{e,1}^2 = \infty$. 
17. Formulas for the Simplified Case and Numerical Example

- **Resulting formulas**: \( x_i = \tilde{x}_i - \lambda \), where
  \[
  \lambda \overset{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i - \tilde{X}_1.
  \]

- **Case study**: \( n = 4 \) cells,
  - with the high spatial resolution accuracy \( \sigma_h = 0.5 \)
  - and the high spatial resolution estimates (in each of these cells)
    \[
    \tilde{x}_1 = 2.0, \quad \tilde{x}_2 = 3.0, \quad \tilde{x}_3 = 5.0, \quad \tilde{x}_4 = 6.0;
    \]
  - the corresponding low spatial resolution estimate is \( \tilde{X}_1 = 3.7 \).
18. Estimates of High and Low Spatial Resolution: Illustration

<table>
<thead>
<tr>
<th>( \tilde{x}_1 )</th>
<th>( \tilde{x}_2 )</th>
<th>( \tilde{x}_3 )</th>
<th>( \tilde{x}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>3.0</td>
<td>5.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

\[ \tilde{X}_1 = 3.7 \]
19. Numerical Example: Discussion

- We assume that the low spatial resolution estimate is accurate ($\sigma_l \approx 0$).

- So, the average of the four cell values is equal to the result $\tilde{X}_1 = 3.7$ of this estimate:

$$\frac{x_1 + x_2 + x_3 + x_4}{4} \approx 3.7.$$ 

- For the high spatial resolution estimates $\tilde{x}_i$, the average is slightly different:

$$\frac{\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 + \tilde{x}_4}{4} = \frac{2.0 + 3.0 + 5.0 + 6.0}{4} = 4.0 \neq 3.7.$$ 

- *Reason:* high spatial resolution estimates are much less accurate: $\sigma_h = 0.5$.

- We use the low spatial resolution estimate to “correct” the high spatial resolution estimate.
20. Numerical Example: Results

- Here, the correcting term takes the form
  \[ \lambda = \frac{\tilde{x}_1 + \ldots + \tilde{x}_n}{n} - \tilde{X}_1 = \frac{2.0 + 3.0 + 5.0 + 6.0}{4} - 3.7 = 4.0 - 3.7 = 0.3. \]

- So, the corrected ("fused") values \( x_i \) take the form:
  \[ x_1 = \tilde{x}_1 - \lambda = 2.0 - 0.3 = 1.7; \quad x_2 = \tilde{x}_2 - \lambda = 3.0 - 0.3 = 2.7; \]
  \[ x_3 = \tilde{x}_3 - \lambda = 5.0 - 0.3 = 4.7; \quad x_4 = \tilde{x}_4 - \lambda = 6.0 - 0.3 = 5.7. \]

- For these corrected values, the arithmetic average is equal to the low spatial resolution estimate:
  \[ \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{1.7 + 2.7 + 4.7 + 5.7}{4} = 3.7. \]
21. The Result of Model Fusion: Simplified Setting

\[
\begin{array}{|c|c|}
\hline
\tilde{x}_1 & 1.7 \\
\hline
\tilde{x}_2 & 2.7 \\
\hline
\tilde{x}_3 & 4.7 \\
\tilde{x}_4 & 5.7 \\
\hline
\end{array}
\]
22. Taking $\sigma_{e,j}$ Into Account

- **Idea:** take into account the requirement that
  - the actual values in each cell are approximately equal to $\tilde{X}_1$,
  - with the accuracy $\sigma_{e,1}$ equal to the empirical standard deviation.

- **Resulting formulas:** $\mu = \frac{1}{n} \cdot \frac{\lambda}{\sigma_h^2} = \frac{1}{n} \cdot \frac{1}{\sigma_h^2}$, and
  
  $$x_i = \frac{\tilde{x}_i - \lambda}{\frac{\sigma_h^2}{\sigma_{e,1}^2} + \frac{\sigma_h^2}{\sigma_{e,1}^2}} + \frac{\sigma_h^2}{\sigma_{e,1}^2} \cdot \frac{\frac{\sigma_h^2}{\sigma_{e,1}^2}}{1 + \frac{\sigma_h^2}{\sigma_{e,1}^2}}.$$

23. Taking $\sigma_{e,j}$ Into Account: Numerical Example

- **General idea:** the actual values in each cell are approximately equal to $\tilde{X}_1$.

- **In our example:** $x_i \approx \tilde{X}_1$, with the accuracy

  $$\sigma^2_{e,1} = \frac{1}{4} \cdot \sum_{i=1}^{4} (\tilde{x}_i - E_1)^2,$$
  where $E_1 = \frac{1}{4} \cdot \sum_{i=1}^{4} \tilde{x}_i$.

- Here, $E_1 = \frac{1}{4} \cdot \sum_{i=1}^{4} \tilde{x}_i = \frac{\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 + \tilde{x}_4}{4} = 4.0$, thus,

  $$\sigma^2_{e,1} = \frac{(2.0 - 4.0)^2 + (3.0 - 4.0)^2 + (5.0 - 4.0)^2 + (6.0 - 4.0)^2}{4} = \frac{4 + 1 + 1 + 4}{4} = \frac{10}{4} = 2.5.$$ 

- Hence $\sigma_{e,1} \approx 1.58$. 
24. Taking $\sigma_{e,j}$ Into Account (cont-d)

- **Reminder:** $x_i = \frac{1}{1 + \frac{\sigma^2}{\sigma^2_{e,1}}} \cdot (\tilde{x}_i - \lambda) + \frac{\sigma^2}{\sigma^2_{e,1}} \cdot \tilde{X}_1.$

- Here, $\sigma_h = 0.5$, $\sigma^2_{e,1} = 2.5$, $\frac{\sigma^2}{\sigma^2_{e,1}} = \frac{0.25}{2.5} = 0.1$, so

\[
\frac{1}{1 + \frac{\sigma^2}{\sigma^2_{e,1}}} = \frac{1}{1.1} \approx 0.91, \quad \text{and} \quad \frac{\sigma^2}{\sigma^2_{e,1}} \cdot \tilde{X}_1 = \frac{0.1}{1.1} \cdot 3.7 \approx 0.34;
\]

\[
x_1 \approx 0.91 \cdot (2.0 - 0.3) + 0.34 \approx 1.89;
x_2 \approx 0.91 \cdot (3.0 - 0.3) + 0.34 \approx 2.79;
x_3 \approx 0.91 \cdot (5.0 - 0.3) + 0.34 \approx 4.62;
x_4 \approx 0.91 \cdot (6.0 - 0.3) + 0.34 \approx 5.53.
\]
25. The Result of Model Fusion: General Setting

\[ \tilde{x}_1 \approx 1.89 \quad \tilde{x}_2 \approx 2.79 \]
\[ \tilde{x}_3 \approx 4.62 \quad \tilde{x}_4 \approx 5.53 \]

- The arithmetic average of these four values is equal to
  \[ \frac{x_1 + x_2 + x_3 + x_4}{4} \approx \frac{1.89 + 2.79 + 4.62 + 5.53}{4} \approx 3.71. \]

- So, within our computation accuracy, it coincides with the low spatial resolution estimate \( \bar{X}_1 = 3.7 \).
26. Model Fusion: Case of Interval Uncertainty

• We take into account three different types of approximate equalities:
  - Each high spatial resolution estimate $\tilde{x}_i$ is approximately equal to the actual value $x_i$:
    $$\tilde{x}_i - \Delta_{h,i} \leq x_i \leq \tilde{x}_i + \Delta_{h,i}.$$  
  - Each lower spatial resolution value $\tilde{X}_j$ is $\approx$ to the average of values of all the cells $x_i(1,j), \ldots, x_i(k_j,j)$:
    $$\tilde{X}_j - \Delta_{l,j} \leq \sum_i w_{j,i} \cdot x_i \leq \tilde{X}_j + \Delta_{l,j}.$$  
  - Finally, we have prior bounds $x_{pr,i}$ and $\bar{x}_{pr,i}$ on the values $x_i$, i.e., bounds for which
    $$x_{pr,i} \leq x_i \leq \bar{x}_{pr,i}.$$  

• Our objective is to find, for each $k = 1, \ldots, n$, the range $[\underline{x}_k, \overline{x}_k]$ of possible values of $x_k$. 
27. Additional Results

- Additional problem: need to fuse discrete and continuous data
- Auxiliary problem: estimating accuracy of fused models
28. Additional Problem: Need to Fuse Discrete and Continuous Models

- Traditionally, seismic models are *continuous*: the velocity smoothly changes with depth.
- In contrast, the gravity models are *discrete*: we have layers, in each of which the velocity is constant.
- The abrupt transition corresponds to a steep change in the continuous model.
- Both models locate the transition only approximately.
- So, if we simply combine the corresponding values value-by-value, we will have *two* transitions instead of one:
  - one transition where the continuous model has it, and
  - another transition nearby where the discrete model has it.
29. What We Plan to Do

- *We want* to avoid the misleading double-transition models.

- *Idea:* first fuse the corresponding transition locations.

- *In this paper,* we provide an algorithm for such location fusion.

- *Specifically,* first, we formulate the problem in the probabilistic terms.

- *Then,* we provide an algorithm that produces the most probable transition location.

- *We show* that the result of the probabilistic location algorithm is in good accordance with common sense.

- *We also show* how the commonsense intuition can be reformulated in fuzzy terms.
30. Available Data: What is Known and What Needs to Be Determined

- For each location, in the discrete model, we have the exact depth \( z_d \) of the transition.
- In contrast, for the continuous model, we do not have the abrupt transition.
- Instead, we have velocity values \( v(z) \) at different depths.
- We must therefore extract the corresponding transition value \( z_c \) from the velocity values.
- To be more precise, we have values \( v_1, v_2, \ldots, v_i, \ldots, v_n \) corresponding to different depths.
- We need to find \( i \) for which the transition occurs between the depths \( i \) and \( i + 1 \).
31. Probabilistic Approach

- The difference \( \Delta v_j = v_j - v_{j+1} \) \( (j \neq i) \) is caused by many independent factors.

- Due to the Central Limit Theorem, we thus assume that it is normally distributed, with probability density

\[
p_j \overset{\text{def}}{=} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \exp \left( -\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2 \right).
\]

- The value \( \Delta v_i \) at the transition depth \( i \) is not described by the normal distribution.

- We assume that differences corresponding to different depths \( j \) are independent, so:

\[
L_i = \prod_{j \neq i} p_j = \prod_{j \neq i} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \exp \left( -\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2 \right).
\]
32. How to Find the Location: The General Idea of the Maximum Likelihood Approach

- **Reminder**: the likelihood of each model is:
  \[ L_i = \prod_{j \neq i} p_j = \prod_{j \neq i} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \exp \left( -\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2 \right) . \]

- **Natural idea**: select the parameters for which the likelihood of the observed data is the largest.

- The value \( L_i \) is the largest if and only if \( -\ln(L_i) \) is the smallest: 
  \[ -\ln(L_i) = \text{const} + \frac{1}{2 \cdot \sigma^2} \cdot \sum_{j \neq i} (\Delta v_j)^2 \rightarrow \min . \]

- This sum is equal to 
  \[ \sum_{j \neq i} (\Delta v_j)^2 = \sum_{j=1}^{n-1} (\Delta v_j)^2 - (\Delta v_i)^2 . \]

- The first term in this expression does not depend on \( i \).

- Thus, the difference is the smallest \( \Leftrightarrow \) the value \( (\Delta v_i)^2 \) is the largest \( \Leftrightarrow |\Delta v_i| \) is the largest.
33. Resulting Location

- **We want**: to select the most probable location of the transition point.

- **We select**: the depth $i_0$ for which the absolute value $|\Delta v_i|$ of the difference $\Delta v_i = v_{i+1} - v_i$ is the largest.

- This conclusion seems to be very reasonable:
  - the most probable location of the actual abrupt transition between the layers
  - is the depth at which the measured difference is the largest.
34. The Results of the Probabilistic Approach are in Good Accordance with Common Sense

- Intuitively, for each depth $i$, our confidence that $i$ is a transition point depends on the difference $|\Delta v_i|$: 
  - the smaller the difference, the less confident we are that this is the actual transition depth, and
  - the larger the difference, the more confident we are that this is the actual transition depth.

- In our probabilistic model, we select a location with the largest possible value $|\Delta v_i|$.

- This shows that the probabilistic model is in good accordace with common sense.

- This coincidence increases our confidence in this result.
35. It May Be Useful to Formulate the Common Sense Description in Fuzzy Terms

- Fuzzy logic is known to be a useful way to formalize imprecise commonsense reasoning.
- Common sense: the degree of confidence $d_i$ that $i$ is a transition point is $f(|\Delta v_i|)$, for some monotonic $f(z)$.
- It is reasonable to select a value $i$ for which our degree of confidence is the largest $d_i = f(|\Delta v_i|) \to \max$.
- Since $f(z)$ is increasing, this is equivalent to $|\Delta v_i| \to \max$.
- Of course, to come up with this conclusion, we do not need to use the fuzzy logic techniques.
- However, this description may be useful if we also have other expert information.
36. How Accurate Is This Location Estimate?

- **Reminder:** the likelihood has the form
  \[ L_i = \prod_{j \neq i} p_j = \prod_{j \neq i} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \exp \left( -\frac{1}{2 \cdot \sigma^2} \cdot (\Delta v_j)^2 \right). \]

- **We have found** the most probable transition \( i_0 \) as the value for which \( L_i \) is the largest.

- **Similarly:** we can find \( \sigma \) for which \( L_i \) is the largest:
  \[ \sigma^2 = \frac{1}{n - 2} \cdot \sum_{j \neq i_0} (\Delta v_j)^2. \]

- The probability \( P_i \) that the transition is at location \( i \) is proportional to \( L_i \):
  \[ P_i = c \cdot L_i. \]

- The coefficient \( c \) can be determined from the condition that the total probability is 1:
  \[ 1 = \sum_i P_i = c \cdot \sum_{i=1}^n L_i. \]

- So, \( c = (\sum L_i)^{-1}. \)
37. Accuracy of the Location Estimate (cont-d)

- The mean square deviation $\sigma_0^2$ of the actual transition depth from our estimate $i_0$ is defined as

$$\sigma_0^2 = \sum_{i=1}^{n-1} (i - i_0)^2 \cdot P_i.$$ 

- We know that $P_i = c \cdot L_i$, and we have formulas for computing $L_i$ and $c$, so we can compute $\sigma_0$.

- We applied this algorithm to the seismic model of El Paso area, and got $\sigma_0 \approx 1.5$ km.

- This value is of the same order (1-2 km) as the difference between:
  - the border depth estimates coming from the seismic data and
  - the border depth coming from the gravity data.
38. **How toFuse the Depth Estimates**

- Now, we have two estimates for the transition depth:
  - the estimate $i_d$ from the discrete (gravity) model;
  - the estimate $i_0$ from the continuous (seismic) model.
- The estimate $i_d$ comes from a standard statistical analysis, so we know standard deviation $\sigma_d$.
- For $i_0$, we already know the standard deviation $\sigma_0$.
- It is reasonable to assume that both differences $i_d - i$ and $i_0 - i$ are normally distributed and independent:
  $$p_i = \exp\left(-\frac{(i_d - i_f)^2}{2 \cdot \sigma_d^2}\right) \cdot \exp\left(-\frac{(i_0 - i_f)^2}{2 \cdot \sigma_0^2}\right).$$
- The most probable location $i$ is when $p_i \rightarrow \max$, i.e.:
  $$i_f = \frac{i_d \cdot \sigma_d^{-2} + i_0 \cdot \sigma_0^{-2}}{\sigma_d^{-2} + \sigma_0^{-2}}.$$
39. Towards Fusing Actual Maps

- In the discrete model:
  - values $i < i_d$ correspond to the upper zone;
  - values $i > i_d$ correspond to the lower zone.

- Similarly, in the continuous model:
  - values $i < i_0$ correspond to the upper zone;
  - values $i > i_0$ correspond to the lower zone.

- So, for depths $i \leq \min(i_0, i_d)$ and $i \geq \max(i_0, i_d)$, both models correctly describe the zone.

- For these depths, we can simply fuse the values from both models.

- We can fuse them similarly to how we fused the depths.

- For intermediate depths, we need to adjust the models: e.g., by taking the nearest value from the correct zone.
40. How to Fuse the Actual Maps: First Stage

- **First**: we adjust both models so that they both have a transition at depth $i_f$.
- **Adjusting the discrete model** is easy: we replace
  - the original depth $i_d$
  - with the new (more accurate) fused value $i_f$.
- **Adjusting the continuous model**:
  - when $i_f < i_0$, the values at depths $i$ between $i_f$ and $i_0$ are erroneously assigned to the upper zone;
  - these values $v_i$ must be replaced by the value of the nearest point at the lower zone $v_{i_0+1}$;
  - when $i_f > i_0$, the values at depths $i$ between $i_0$ and $i_f$ are erroneously assigned to the lower zone;
  - these values $v_i$ must be replaced by the value of the nearest point at the upper zone $v_{i_0}$. 
41. How to Merge the Adjusted Models

- For each depth $i$, we now have two adjusted values $v'_i$ and $v''_i$ corresponding to two adjusted models.
- Let $\sigma'$ and $\sigma''$ be the corresponding standard deviations.
- It is reasonable to assume that both differences $v'_i - v_i$ and $v''_i - v_i$ are normally distributed and independent:

$$p(v_i) = \exp\left(-\frac{(v'_i - v_i)^2}{2 \cdot (\sigma')^2}\right) \cdot \exp\left(-\frac{(v''_i - v_i)^2}{2 \cdot (\sigma'')^2}\right).$$

- The most probable value $\tilde{v}_i$ is when $p(v_i) \to \max$, i.e.:

$$\tilde{v}_i = \frac{v'_i \cdot (\sigma')^{-2} + v''_i \cdot (\sigma'')^{-2}}{(\sigma')^{-2} + (\sigma'')^{-2}}.$$
42. Auxiliary Problem: How to Estimate Accuracy of Fused Models

- *Calibration* is possible when we have a “standard” (several times more accurate) measuring instrument (MI).
- In geophysics, seismic (and other) methods are state-of-the-art.
- No method leads to more accurate determination of the densities.
- In some practical situations, we can use two similar MIs to measure the same quantities $x_i$.
- In geophysics, we want to estimate the accuracy of a model, e.g., a seismic model, a gravity-based model.
- In this situation, we do not have two similar applications of the same model.
43. Maximum Likelihood (ML) Approach Cannot Be Applied to Estimate Model Accuracy

- We have several quantities with (unknown) actual values \( x_1, \ldots, x_i, \ldots, x_n \).
- We have several measuring instruments (or geophysical methods) with (unknown) accuracies \( \sigma_1, \ldots, \sigma_m \).
- We know the results \( x_{ij} \) of measuring the \( i \)-th quantity \( x_i \) by using the \( j \)-th measuring instrument.
- At first glance, a reasonable idea is to find all the unknown quantities \( x_i \) and \( \sigma_j \) from ML:

\[
L = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi} \cdot \sigma_j} \cdot \exp \left( - \frac{(x_{ij} - x_i)^2}{2\sigma_j^2} \right) \to \max.
\]

- Fact: the largest value \( L = \infty \) is attained when, for some \( j_0 \), we have \( \sigma_{j_0} = 0 \) and \( x_i = x_{ij_0} \) for all \( i \).
- Problem: this is not physically reasonable.
44. How to Estimate Model Accuracy: Idea

- For every two models, the difference $x_{ij} - x_{ik} = \Delta x_{ij} - \Delta x_{ik}$ is normally distributed, w/variance $\sigma_j^2 + \sigma_k^2$.
- We can thus estimate $\sigma_j^2 + \sigma_k^2$ as
  \[
  \sigma_j^2 + \sigma_k^2 \approx A_{jk} \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - x_{ik})^2.
  \]
- So, $\sigma_1^2 + \sigma_2^2 \approx A_{12}$, $\sigma_1^2 + \sigma_3^2 \approx A_{13}$, and $\sigma_2^2 + \sigma_3^2 \approx A_{23}$.
- By adding all three equalities and dividing the result by two, we get
  \[
  \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = \frac{A_{12} + A_{13} + A_{23}}{2}.
  \]
- Subtracting, from this formula, the expression for $\sigma_2^2 + \sigma_3^2$, we get
  \[
  \sigma_1^2 \approx \frac{A_{12} + A_{13} - A_{23}}{2}.
  \]
- Similarly, $\sigma_2^2 \approx \frac{A_{12} + A_{23} - A_{13}}{2}$ and $\sigma_3^2 \approx \frac{A_{13} + A_{23} - A_{12}}{2}$.
45. How to Estimate Model Accuracy: General Case and Challenge

- **General case**: we may have $M \geq 3$ different models.
- Then, we have \[
\frac{M \cdot (M - 1)}{2}\]
different equations \[\sigma_j^2 + \sigma_k^2 \approx A_{jk}\] to determine $M$ unknowns $\sigma_j^2$.
- When $M > 3$, we have more equations than unknowns,
- So, we can use the Least Squares method to estimate the desired values $\sigma_j^2$.
- **Challenge**: the formulas $\sigma_1^2 \approx \tilde{V}_1 \overset{\text{def}}{=} \frac{A_{12} + A_{13} - A_{23}}{2}$ are approximate.
- Sometimes, these formulas lead to physically meaningless negative values $\tilde{V}_1$.
- It is therefore necessary to modify the above formulas, to avoid negative values.
46. An Idea of How to Deal With This Challenge

• The negativity challenge is caused by the fact that the estimates \( \tilde{V}_j \) for \( \sigma_j^2 \) are approximate.

• For large \( n \), the difference \( \Delta V_j \stackrel{\text{def}}{=} \tilde{V}_j - \sigma_j^2 \) is asymptotically normally distributed, with asympt. 0 mean.

• We can estimate the standard deviation \( \Delta_j \) for this difference.

• Thus, \( \sigma_j^2 = \tilde{V}_j - \Delta V_j \) is normally distributed with mean \( \tilde{V}_j \) and standard deviation \( \Delta_j \).

• We also know that \( \sigma_j^2 \geq 0 \).

• As an estimate for \( \sigma_j^2 \), it is therefore reasonable to use a conditional expected value \( E \left( \tilde{V}_j - \Delta V_j \mid \tilde{V}_j - \Delta V_j \geq 0 \right) \).

• This new estimate is an expected value of a non-negative number and thus, cannot be negative.
47. Resulting Algorithm

- **Input**: for each value \( x_i \) (\( i = 1, \ldots, n \)), we have three estimates \( x_{i1}, x_{i2}, \text{ and } x_{i3} \) corr. to three diff. models.

- **Objective**: to estimate the accuracies \( \sigma_j^2 \) of these three models.

- First, for each \( j \neq k \), we compute

\[
A_{jk} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - x_{ik})^2.
\]

- Then, we compute

\[
\tilde{V}_1 = \frac{A_{12} + A_{13} - A_{23}}{2}; \quad \tilde{V}_2 = \frac{A_{12} + A_{23} - A_{13}}{2};
\]

\[
\tilde{V}_3 = \frac{A_{13} + A_{23} - A_{12}}{2}.
\]

- After that, for each \( j \), we compute

\[
\Delta_j^2 = \frac{1}{n} \left( \left( \tilde{V}_j \right)^2 + \tilde{V}_j \cdot \tilde{V}_k + \tilde{V}_j \cdot \tilde{V}_\ell + \tilde{V}_k \cdot \tilde{V}_\ell \right).
\]
48. Resulting Algorithm (cont-d)

- **Reminder:** we compute \( \tilde{V}_j = \frac{A_{jk} + A_{j\ell} - A_{kl}}{2} \) and
  \[
  \Delta_j^2 = \frac{1}{n} \cdot \left( \left( \tilde{V}_j \right)^2 + \tilde{V}_j \cdot \tilde{V}_k + \tilde{V}_j \cdot \tilde{V}_\ell + \tilde{V}_k \cdot \tilde{V}_\ell \right).
  \]

- Then, we compute the auxiliary ratios \( \delta_j = \frac{\tilde{V}_j}{\Delta_j} \).

- Finally, we return as an estimate \( \tilde{\sigma}_j^2 \) for \( \sigma_j^2 \), the value
  \[
  \tilde{\sigma}_j^2 = \tilde{V}_j + \frac{\Delta_j}{\sqrt{2\pi}} \cdot \Phi\left(-\frac{\delta_j^2}{2}\right).
  \]

- These non-negative estimates \( \tilde{\sigma}_j^2 \) can now be used to fuse the models: for each \( i \), we take \( x_i = \frac{\sum \tilde{\sigma}_j^{-2} \cdot x_{ij}}{\sum \tilde{\sigma}_j^{-2}} \).
49. Conclusions

- In many practical situations, there is a need to combine (fuse) data from different datasets.
- Ideal approach of *joint inversion* – which uses all the data from all the datasets – is often not yet practical.
- Main idea of *model fusion*: process each dataset separately and fuse the resulting models.
- In this thesis, algorithms are proposed for fusing models with different accuracy and spatial resolution.
- This thesis also addresses additional challenge:
  - fusing discrete and continuous models;
  - estimating the accuracy of fused models.
- This work can help geophysicists combine complementary models.
50. Acknowledgments

- This work was supported by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence).

- The author is greatly thankful:
  - to Drs. Ann Gates, Vladik Kreinovich, and Aaron Velasco for their help and support, and
  - to family and friends for being there with me.